SIAM J. CONTROL AND OPTIMIZATION
Vol. 33, No. 1, pp. 185–207, January 1995

009

NONLINEAR H^∞ OPTIMIZATION: A CAUSAL POWER SERIES APPROACH*

CIPRIAN FOIAS†, CAIXING GU†, AND ALLEN TANNENBAUM‡

Abstract. In this paper, using a power series methodology a design procedure applicable to analytic nonlinear plants is described. The technique used is a generalization of the linear H^∞ theory. In contrast to previous work on this topic ([*Indiana J. Math.*, 36 (1987), pp. 693–709], [*Oper. Theory Adv. Appl.*, 41 (1989), pp. 255–277], [*SIAM J. Control Optim.*, 27 (1989), pp. 842–860]), the authors are now able to incorporate explicitly a causality constraint into the theory. In fact, it is shown that it is possible to reduce a causal optimal design problem (for nonlinear systems) to a classical interpolation problem solvable by the commutant lifting theorem [*Harmonic Analysis of Operators on Hilbert Space*, North-Holland, Amsterdam, 1970], [*The Commutant Lifitng Approach to Interpolation Problems*, Birkhäuser, Boston, 1990].

Key words. nonlinear systems, H^∞ optimization, causality, commutant lifting theorem, interpolation theory, Volterra series

AMS subject classifications. primary 47A20, secondary 47A99, 93B35, 93C05

1. Introduction. In this paper, we continue our work on finding a suitable, implementable nonlinear extension of the powerful linear H^∞ design methodology. In what follows, we will just consider discrete-time systems, even though the techniques elucidated below carry over to the continuous-time setting as well.

Our approach is based on previous work ([14], [15]) in which we considered systems described by analytic input/output operators. A key idea here involved the expression of each n-linear term of a suitable Taylor expansion of the given operator as an equivalent linear operator acting on a certain associated tensor space, which allowed us to iteratively apply the classical commutant lifting theorem in designing a compensator. (Our class of operators includes Volterra series [9].)

More precisely, in such an approach we are reduced to applying the classical (linear) commutant lifting theorem to an H^2-space defined on some D^n (where D denotes the unit disc). Now when we apply the classical result to D^n ($n \geq 2$), even though time-invariance is preserved (that is, commutation with the appropriate shift), causality may be lost. Indeed, for systems described by analytic functions on the disc D (these correspond to stable, discrete-time, one-dimensional (1D) systems), time-invariance (that is, commutation with the unilateral shift) implies causality. For analytic functions on the n-disc ($n > 1$), this is not necessarily the case. For dynamical system control design and for any physical application, this is, of course, a major drawback for such an approach. (The compensators we obtained were "weakly causal" and causal approximations were discussed.)

Hence for a dilation result in $H^2(D^n)$ we must include the causality constraint explicitly in the set-up of the dilation problem. It is precisely this problem that motivated the mathematical operator-theoretic work of [16] and [13], which incorporated Arveson theory [1] into the dilation, commutant lifting framework.

* Received by the editors August 24, 1992; accepted for publicaiton (in revised form) August 24, 1993. This work was supported in part by a grant from the Research Fund of Indiana University, and by National Science Foundation grants DMS-8811084 and ECS-9122106, Air Force Office of Scientific Research grants AFOSR-90-0024 and F49620-94-1-00S8DEF, Army Research Office grants DAAL03-91-G-0019, DAAH04-94-G-0054, and DAAH04-93-G-0032, and by Image Evolutions, Ltd.

† Department of Mathematics, Indiana University, Bloomington, Indiana 47405.

‡ Department of Electrical Engineering, University of Minnesota, Minneapolis, Minnesota 55455.

While the general method explicated in this paper is based on a causal extension of the commutant lifting theorem, for the purposes of the operators and spaces that appear in control we will give a direct simple method for finding the optimal causal compensators. In fact, we will show that *the computation of an optimal causal nonlinear compensator may be reduced to a classical interpolation problem.*

We now briefly outline the contents of this paper. In §2, we define *causality* and *time-invariance* as applied to analytic mappings. We show in particular that while in the linear case, time-invariance and boundedness imply causality, this is not true in general in the nonlinear setting. In §3, we formulate the causal optimization problem to be studied. In §4, we discuss the Fourier representation of certain Hilbert spaces, a technique that we apply throughout the paper. In §5, we prove the main theoretical result of this paper in which we show how to reduce a causal optimization problem to a problem solvable via the classical commutant lifting theorem [25]. This is summarized in a computational algorithm in §§6 and 7. Sections 8 and 9 are then concerned with our formulation of the nonlinear generalization of the H^∞ sensitivity minimization problem, which is then solved via a *causal iterative commutant lifting* method in §10. Section 11 is devoted to a natural control interpretation of our optimization procedure, while §12 is connected to computational aspects of our work, namely a nonlinear notion of *rationality* that reduces our work to finite-dimensional skew Toeplitz calculations. We illustrate our methods with an example in §13, and finally in §14, we make some concluding remarks.

We conclude this section by noting that there have been other approaches to nonlinear H^∞. These include a nonlinear commutant lifting theorem [3], [4], and a very promising nonlinear game-theoretic approach [7] as well as a nonlinear version of Ball–Helton theory [6], and the recent work in [26].

Once again, we will just consider discrete-time systems in what follows.

2. Causal analytic mappings. In this section, we will define the class of nonlinear input/output operators that we will study in this paper. To do this, we will first need to discuss a few standard results about analytic mappings on Hilbert spaces. See [3], [4], [14], [15], [21] and the references therein for complete details.

Let $\mathcal{G}$ and $\mathcal{H}$ denote complex separable Hilbert spaces. Set

$$B_{r_o}(\mathcal{G}) := \{g \in \mathcal{G} : \|g\| < r_o\}$$

(the open ball of radius r_o in $\mathcal{G}$ about the origin). Then we say that a mapping $\phi : B_{r_o}(\mathcal{G}) \to \mathcal{H}$ is *analytic* if the complex function $(z_1, \ldots, z_n) \mapsto \langle \phi(z_1 g_1 + \ldots + z_n g_n), h\rangle$ is analytic in a neighborhood of $(1, 1, \ldots, 1) \in \mathbf{C}^n$ as a function of the complex variables $z_1, \ldots, z_n$ for all $g_1, \ldots, g_n \in \mathcal{G}$ such that $\|g_1 + \ldots + g_n\| < r_o$, for all $h \in \mathcal{H}$, and for all $n > 0$.

We will now assume that $\phi(0) = 0$. It is easy to see that if $\phi : B_{r_o}(\mathcal{G}) \to \mathcal{H}$ is analytic, then ϕ admits a convergent Taylor series expansion ([21, p. 97]), i.e.,

$$\phi(g) = \phi_1(g) + \phi_2(g, g) + \cdots + \phi_n(g, \cdots, g) + \cdots,$$

where $\phi_n : \mathcal{G} \times \cdots \times \mathcal{G} \to \mathcal{H}$ is an n-linear map. Clearly, without loss of generality we may assume that the n-linear map $(g_1, \cdots, g_n) \to \phi(g_1, \ldots, g_n)$ is symmetric in the arguments $g_1, \ldots, g_n$. This assumption will be made throughout this paper for the various analytic maps which we consider. For ϕ a Volterra series, ϕ_n is basically the nth-Volterra kernel.

Now set

$$\hat{\phi}_n(g_1 \otimes \cdots \otimes g_n) := \phi_n(g_1, \ldots, g_n).$$

Then $\hat{\phi}_n$ extends in a unique manner to a dense set of $\mathcal{G}^{\otimes n} := \mathcal{G} \otimes \cdots \otimes \mathcal{G}$ (tensor product taken n times). Note by $\mathcal{G}^{\otimes n}$ we mean the Hilbert space completion of the algebraic tensor product of the $\mathcal{G}$'s. Clearly if $\hat{\phi}_n$ has finite norm on this dense set, then $\hat{\phi}_n$ extends by continuity to a bounded linear operator $\hat{\phi}_n : \mathcal{G}^{\otimes n} \to \mathcal{H}$. By abuse of notation, we will set $\phi_n := \hat{\phi}_n$. (Recall that an n-linear map on $G \times G \times \cdots \times G$ (product taken n times) becomes linear on the tensor product $\mathcal{G}^{\otimes n}$. For details about the construction of the tensor product, see [2, pp. 24–27].)

We now recall the following standard definitions.

DEFINITION 1. (i) *Notation as above. By a* majorizing sequence *for the analytic map ϕ, we mean a positive sequence of numbers α_n $n = 1, 2, \ldots$ such that $\|\phi_n\| < \alpha_n$ for $n \geq 1$. Suppose that $\rho := \limsup {\alpha_n}^{1/n} < \infty$. Then it is completely standard that the Taylor series expansion of ϕ converges at least on the ball $B_r(\mathcal{G})$ of radius $r = 1/\rho$* ([21, p. 97]).

(ii) *If ϕ admits a majorizing sequence as in* (i), *then we will say that ϕ is* majorizable.

Let $H^2_K(D^n)$ denote the standard Hardy space of $\mathbf{C}^K$-valued analytic functions on the n-disc D^n (D denotes the unit disc) with square integrable boundary values. We set $H^2_K := H^2_K(D)$ and and $H^2 := H^2_1$. We denote the shift on $H^2_K(D^n)$ by $S_{(n)}$. Note that $S_{(n)}$ is defined by multiplication by the function $(z_1 \cdots z_n)$. On H^2_K we set $S_{(1)} =: U$ (U is given by multiplication by z).

We now consider an analytic map ϕ with $\mathcal{G} = \mathcal{H} = H^2_k$. Note that

$$H^2_k \otimes \cdots \otimes H^2_k = (H^2_k)^{\otimes n} \cong H^2_K(D^n) \quad \text{with } K = k^n, \tag{1}$$

where we map $1 \otimes \cdots \otimes z \otimes \cdots \otimes 1$ (z in the ith place) to z_i, $i = 1, \ldots, n$. Clearly, $S_{(n)}$ corresponds to $U^{\otimes n}$ under this identification.

We will identify ϕ_n as a bounded linear map from $H^2_K(D^n) \to H^2_k$ via the canonical isomorphism (1). Then we say that ϕ is *time-invariant* if

$$\phi_n S_{(n)} = U \phi_n \quad \forall n \geq 1. \tag{2}$$

(We will also say each ϕ_n is *time-invariant.*) Equivalently, this means that $U\phi = \phi \circ U$ on some open ball about the origin in which ϕ is defined.

Now set

$$P^{(j)}_{(n)} := I - S^j_{(n)} S^{*j}_{(n)}, \qquad j \geq 1, \quad n \geq 1.$$

Note

$$P^{(j)} := P^{(j)}_{(1)} = I - U^j U^{*j}.$$

Then we say that ϕ is *causal* if

$$P^{(j)} \phi_n = P^{(j)} \phi_n P^{(j)}_{(n)}, \qquad j \geq 1, \quad n \geq 1. \tag{3}$$

(We also say each ϕ_n is *causal.*) Equivalently, $\phi_n : H^2_K(D^n) \to H^2_k$ is causal if for $F(z_1, \ldots, z_n) \in H^2_K(D^n)$,

$$F(z_1, \ldots, z_n) = \sum_{i_1, \ldots, i_n \geq 0} F_{i_1, \ldots, i_n} z_1^{i_1} \ldots z_n^{i_n}, \qquad \phi_n(F)(z) := \sum_{m \geq 0} f_m z^m,$$

each f_m only depends on

$$\{F_{i_1,\dots,i_n} : 0 \le i_1, \dots, i_n \le m\}.$$

This means that for

$$F(z_1, \dots, z_n) = \sum_{\max\{i_1,\dots,i_n\} \ge m} F_{i_1,\dots,i_n} z_1^{i_1} \dots z_n^{i_n},$$

we have that

$$(I - U^m U^{*m}) \phi_n(F(z_1, \dots, z_n)) = 0. \tag{4}$$

We would now like to discuss the relationship between time-invariance and causality. For simplicity, we assume $k = 1$, i.e., we work with single-input/single-output (SISO) systems. Let $\phi : H^2 \to H^2$ be linear and time-invariant (i.e., intertwines with the shift). Then it is easy to see that ϕ is causal. Indeed, $\phi U = U\phi$ implies

$$\begin{aligned} U^m U^{*m} \phi U^m U^{*m} &= U^m U^{*m} U^m \phi U^{*m} \\ &= U^m \phi U^{*m} \\ &= \phi U^m U^{*m}, \end{aligned}$$

which immediately implies

$$P^{(m)} \phi P^{(m)} = P^{(m)} \phi \quad \forall m \ge 1,$$

that is, ϕ is causal.

In the nonlinear setting however, time-invariance may not imply causality. As a concrete example, let $\phi_o : (H^2)^{\otimes 2} \to H^2$ be a linear operator such that $U^{\otimes 2}\phi_o = \phi_o U$, defined by

$$(\phi_o(f \otimes g))(z) := \sum_{m=0}^{\infty} (f_{m+1} f_m + f_m g_m + f_m g_{m+1}) z^m,$$

where

$$f(z) = \sum_{m=0}^{\infty} f_m z^m, \qquad g(z) = \sum_{m=0}^{\infty} g_m z^m.$$

Now set

$$\phi(f) := \phi_o(f \otimes f), \qquad f \in H^2.$$

Then ϕ is an analytic, time-invariant map. (In fact ϕ is a homogeneous polynomial of degree 2.) But ϕ is not causal. Indeed,

$$\begin{aligned} (P^{(1)}\phi(f))(z) &= 2f_1 f_0 + f_0^2, \qquad z \in D \\ (P^{(1)}\phi(P^{(1)}_{(2)} f))(z) &= f_0^2, \qquad z \in D. \end{aligned}$$

Thus $P^{(1)}\phi(f) \neq P^{(1)}\phi(P^{(1)}_{(2)} f)$, for example for $f(z) := 1 + z$ for $z \in D$. (Note that under the identification (1), $P^{(1)}_{(2)}$ corresponds to $P^{(1)} \otimes P^{(1)}$.)

3. Causal optimization problem. One of the key techniques in this paper will be to reduce a nonlinear generalization of the H^∞ sensitivity minimization problem to a series of *linear causal optimization problems.* (This will be done in §§8–10 below.) In this section, we will formulate this new causal problem.

As above, we let $S_{(n)}$ denote the unilateral shift on $H^2_K(D^n)$ given by multiplication by $(z_1 \cdots z_n)$. Since $H^2_K(D^n)$ will be fixed in the discussion we will let $S := S_{(n)}$. As above, U will denote the unilateral shift on H^2_k given by multiplication by z, and $\Theta \in H^\infty_{k\times k}$ will be an inner $k \times k$ matrix-valued H^∞ function (i.e., a $k \times k$ inner matrix with entries H^∞ scalar functions). Finally $W : H^2_K(D^n) \to H^2_k$ will denote a causal, time-invariant bounded linear operator (in the sense of (2) and (3) above).

We can now state the *causal H^∞-optimization problem* (COP): Find

$$\sigma := \inf\{\|W - \Theta Q\| : Q : H^2_K(D^n) \to H^2_k,\ Q \text{ causal, time-invariant}\}. \tag{5}$$

Moreover, we want to compute an optimal, causal, time-invariant Q_{opt} such that

$$\sigma = \|W - \Theta Q_{\text{opt}}\|. \tag{6}$$

If we drop the causality constraint the solution to problem (5) is provided by the classical commutant lifting theorem [25]. With the causality constraint, the solution to (COP) is abstractly provided by a causal commutant lifting theorem [16], [13].

In this paper, based on this work we will provide a simple solution to the problem (COP) without directly referring to the operator theoretic results of [16] and [13]. In fact, we will show how to directly reduce the computation of σ to a classical interpolation problem handled by the ordinary commutant lifting theorem, a computational procedure for which was given in [14] and [15]. We will also describe how to get the corresponding optimal parameter Q_{opt}.

Our technique will be based on a *reduction theorem* stated in §5. To formulate this result, we will first discuss the Fourier representation, which we do in the next section.

4. Fourier representation. In what follows we must use the Fourier representation of elements of $H^2_K(D^n)$. We refer the reader to [25] for all the details.

We first precisely define all the relevant spaces. First we denote by

$$\ell^2(H^2_K) := \bigoplus_{i=1}^{\infty} H^2_K,$$

the Hilbert space of all column vectors

$$f(z) = [f_1(z), f_2(z), \ldots, f_n(z), \ldots]', \tag{7}$$

($'$ stands for tranpose) such that

$$\|f\|^2 := \sum_{i=1}^{\infty} \|f_i\|^2, \tag{8}$$

is finite. ($\|\ \|$ is our generic symbol for a Hilbert space norm (2-norm) as well as the induced operator norm. So for example in (8), it stands for the usual norm on H^2_K as well as the associated norm on $\ell^2(H^2_K)$.) Thus $\ell^2(H^2_K)$ is a vector-valued Hardy space. Indeed, if $f(z)$ is given by (7), then we may write

$$f(z) = \sum_{m=0}^{\infty} a_m z^m, \tag{9}$$

where each a_m is an infinite column vector with components in $\mathbf{C}^K$, and

$$a_m = \frac{1}{m!}[f_1^{(m)}(0), \ldots, f_j^{(m)}(0), \ldots]'.$$

Clearly,

$$\|f\|^2 = \sum_{m=0}^{\infty} \|a_m\|^2.$$

Conversely, if $f(z) \in \ell^2(H_K^2)$ is given in the form (9) for

$$a_m = [a_{m1}, \ldots, a_{mj}, \ldots]',$$

then $f(z)$ can be written in the form (7), i.e.,

$$f(z) = [f_1(z), \ldots, f_j(z), \ldots]',$$

where

$$f_j(z) = \sum_{m=0}^{\infty} a_{mj} z^m.$$

In what follows, we will either use representation (7) or (9). The context should always make the meaning clear.

Next we let $S_\Phi : \ell^2(H_K^2) \to \ell^2(H_K^2)$ denote the unilateral shift defined by multiplication by z. Then the *Fourier representation of* $H_K^2(D^n)$ is given by the (linear, bounded) operator

$$\Phi := \Phi_n : H_K^2(D^n) \to \ell^2(H_K^2),$$

which is defined by

$$f(z) := \Phi(F(z_1, z_2, \ldots, z_n))$$

$$(10) \qquad := \sum_{m=0}^{\infty} z^m \begin{bmatrix} F_{m,m,\ldots,m} \\ F_{m,\ldots,m,m+1} \\ F_{m,\ldots m+1,m+1} \\ \vdots \\ F_{m+i_1,m+i_2,\ldots,m+i_n} \\ \vdots \end{bmatrix},$$

where

$$F(z_1, \ldots, z_n) = \sum_{j_1,\ldots,j_n \geq 0} F_{j_1,\ldots,j_n} z_1^{j_1} \cdots z_n^{j_n},$$

and $(i_1, \ldots, i_n) \in I_n$ for

$$(11) \qquad I_n := \{(i_1, \ldots, i_n) : i_1, \ldots, i_n \geq 0, \ \min\{i_1, \ldots, i_n\} = 0\}.$$

We order the set I_n in the following manner. We have $(i_1, \ldots, i_n) < (i'_1, \ldots, i'_n)$, if $\max\{i_1, \ldots, i_n\} < \max\{i'_1, \ldots, i'_n\}$. Thus

$$I_n = \bigcup_{k \geq 0} I_n^{(k)},$$

where

$$I_n^{(k)} := \{(i_1, \ldots, i_n) \in I_n : \max\{i_1, \ldots, i_n\} = k\}.$$

Each $I_n^{(k)}$ is then ordered by the lexicographical order.

Note that we are taking $f(z)$ in the form (9) in the above representation. Moreover, note that

$$H_K^2(D^n) = \left\{ F(z_1, \ldots, z_n) = \sum_{j_1,\ldots,j_n \geq 0} F_{j_1,\ldots,j_n} z_1^{j_1} \cdots z_n^{j_n} : \sum_{j_1,\ldots,j_n \geq 0} \|F_{j_1,\ldots,j_n}\|^2 < \infty \right\}.$$

We can also write

$$f(z) = [f_{0,\ldots,0}(z), f_{0,\ldots,1}(z), \ldots, f_{i_1,\ldots,i_n}(z), \ldots]', \tag{12}$$

where

$$f_{i_1,\ldots,i_n}(z) := \sum_{m=0}^{\infty} F_{i_1+m,\ldots,i_n+m} z^m, \tag{13}$$

and $(i_1, \ldots, i_n) \in I_n$.

Next, it is easy to see that $\Phi : H_K^2(D^n) \to \ell^2(H_K^2)$ is an isometry. Indeed, using (10), (12), and (13), we have

$$\begin{aligned} \|\Phi(F)\|^2 &= \|f\|^2 \\ &= \sum_{i_1,\ldots,i_n \in I_n} \|f_{i_1,\ldots,i_n}\|^2 \\ &= \sum_{i_1,\ldots,i_n \in I_n} \|F_{i_1+m,\ldots,i_n+m}\|^2 \\ &= \sum_{j_1,\ldots,j_n \geq 0} \|F_{j_1,\ldots,j_n}\|^2 \\ &= \|F\|^2. \end{aligned}$$

A similar computation shows that the adjoint of Φ is also an isometry, so that Φ is an unitary operator. We next show that

$$\Phi S = S_\Phi \Phi. \tag{14}$$

Indeed, we see that

$$\begin{aligned} \Phi S(F) &= \Phi(z_1 \cdots z_n F(z_1, \ldots, z_n)) \\ &= \Phi\Big(\sum_{j_1,\ldots,j_n \geq 0} F_{j_1,\ldots,j_n} z_1^{j_1+1} \cdots z_n^{j_n+1}\Big) \\ &= \sum_{m=0}^{\infty} z^{m+1} \begin{bmatrix} F_{m,\ldots,m} \\ F_{m,\ldots,m,m+1} \\ F_{m,\ldots,m+1,m+1} \\ \vdots \\ F_{m+i_1,m+i_2,\ldots,m+i_n} \\ \vdots \end{bmatrix} \\ &= z\Phi(F) \\ &= S_\Phi \Phi(F). \end{aligned}$$

By (14), we see that if $W : H^2_K(D^n) \to H^2_k$ is such that $WS = UW$, then the operator $W\Phi^* : \ell^2(H^2_K) \to H^2_k$ satisfies

$$(W\Phi^*)S_\Phi = WS\Phi^* = U(W\Phi^*);$$

that is, $W\Phi^*$ intertwines the shifts S_Φ and U. Consequently, it is standard (see, e.g., [12], or [25, p. 277]) that $W\Phi^*$ is represented by a row vector

$$[W_{0,\dots,0}(z), W_{0,\dots,1}(z), \dots, W_{i_1,\dots,i_n}(z), \dots], \tag{15}$$

for $(i_1, \dots, i_n) \in I_n$. Specifically, for any

$$f(z) = [f_{0,\dots,0}(z), f_{0,\dots,1}(z), \dots, f_{i_1,\dots,i_n}(z), \dots]' \in \ell^2(H^2_K),$$

we have

$$(W\Phi^*)f(z) = \sum_{i_1,\dots,i_n \in I_n} W_{i_1,\dots,i_n}(z) f_{i_1,\dots,i_n}(z). \tag{16}$$

We will write that

$$W\Phi^* \cong [W_{0,\dots,0}(z), W_{0,\dots,1}(z), \dots, W_{i_1,\dots,i_n}(z), \dots], \tag{17}$$

in the sense expressed by (15) and (16).

We would like to make this representation a bit more precise now. Note that the action of $W\Phi^*$ is determined by its action on

$$\ker S^*_\Phi = \{a \in \ell^2(H^2_K) : a \text{ is a column vector with components in } \mathbf{C}^K\}.$$

(This follows from the fact that

$$\ell^2(H^2_K) \cong \bigoplus_{j=0}^{\infty} S^j_\Phi(\ker S^*_\Phi),$$

and that $W\Phi^*$ intertwines the shifts S_Φ and U.) Thus we need only to compute the action of W on

$$\Phi^* \ker S^*_\Phi = \left\{ F(z_1, \dots, z_n) \in H^2_K(D^n) : F(z_1, \dots, z_n) = \sum_{i_1,\dots,i_n \in I_n} F_{i_1,\dots,i_n} z_1^{i_1} \cdots z_n^{i_n} \right\}.$$

(See (11) for the definition of I_n.) By linearity,

$$W\left(\sum_{i_1,\dots,i_n \in I_n} F_{i_1,\dots,i_n} z_1^{i_1} \cdots z_n^{i_n} \right) = \sum_{i_1,\dots,i_n \in I_n} F_{i_1,\dots,i_n} W(z_1^{i_1} \cdots z_n^{i_n}).$$

So by (10) and (16) we have

$$W\Phi^* \cong [W_{0,\dots,0}(z), W_{0,\dots,1}(z), \dots, W_{i_1,\dots,i_n}(z), \dots], \tag{18}$$

where

$$W_{i_1,\dots,i_n}(z) = W(z_1^{i_1} \cdots z_n^{i_n}) \qquad (i_1, \dots, i_n) \in I_n. \tag{19}$$

The above discussion used only the time-invariance for W. In the next proposition, we will write down an explicit expression for the row vector of (18) and (19) associated with $W\Phi^*$ in case W is causal.

PROPOSITION 4.1. *Let $W : H^2_K(D^n) \to H^2_k$ be time-invariant. Then W is causal if and only if*

$$W_{i_1,\ldots,i_n}(z) = z^{\max\{i_1,\ldots,i_n\}} W^c_{i_1,\ldots,i_n}(z) \quad \forall (i_1,\ldots,i_n) \in I_n,$$

where $W^c_{i_1,\ldots,i_n}(z) \in H^\infty_{k\times K}$ (the space of $k \times K$ matrix-valued H^∞ functions).

Proof. By definition, for all $(i_1,\ldots,i_n) \in I_n$ with $\max\{i_1,\ldots,i_n\} = m$, and for all $v \in \mathbf{C}^k$, we have by the causality condition (4) that

$$(I - U^m U^{*m}) W\Phi^*(\Phi(z_1^{i_1}\cdots z_n^{i_n} v)) = (I - U^m U^{*m}) W_{i_1,\ldots,i_n}(z) v = 0.$$

Thus

$$W_{i_1,\ldots,i_n}(z) = z^m W^c_{i_1,\ldots,i_n}(z) = z^{\max\{i_1,\ldots,i_n\}} W^c_{i_1,\ldots,i_n}(z) \quad \forall (i_1,\ldots,i_n) \in I_n, \tag{20}$$

for some $W^c_{i_1,\ldots,i_n}(z) \in H^\infty_{k\times K}$, as required. □

By the above discussion (in particular, Proposition 4.1), we see that for W, Θ as in the (COP) problem (5), we have

$$\begin{aligned}
\sigma &= \inf\{\|W - \Theta Q\| : QS = UQ, Q \text{ causal, time-invariant}\}\\
&= \inf\{\|W\Phi^* - \Theta Q\Phi^*\| : (Q\Phi^*)S_\Phi = U(Q\Phi^*), Q \text{ causal, time-invariant}\}\\
&= \inf\{\|W_1 - \Theta Q_1\| : W_1, Q_1 : \ell^2(H^2_K) \to H^2_k,\ W_1 = W\Phi^*,\\
&\quad Q_1 \cong [q_{0,\ldots,0}(z), zq_{0,\ldots,1}(z), \ldots, zq_{1,\ldots,1,0}(z), z^2 q_{0,\ldots,2}(z), \ldots]\}.
\end{aligned}$$

From now on (unless explicitly stated otherwise), we will just work with the operators $W_1, Q_1 : \ell^2(H^2_K) \to H^2_k$. Essentially, via the unitary equivalence Φ, we are identifying the spaces $H^2_K(D^n)$ and $\ell^2(H^2_K)$. In particular, we identify W with $W_1 = W\Phi^*$, and Q with $Q_1 = Q\Phi^*$. For simplicity of notation, we will denote

$$W = W_1, \qquad Q = Q_1.$$

The context should always make the meaning clear.

We now translate the notions of causality and time-invariance for an operator $W : \ell^2(H^2_K) \to H^2_k$. We will say that W is *time-invariant* if $WS_\Phi = UW$, that is,

$$W \cong [W_{0,\ldots,0}(z), W_{0,\ldots,1}(z), \ldots, W_{i_1,\ldots,i_n}(z), \ldots].$$

Moreover, we say that W is *causal* if the operator $W\Phi : H^2_K(D^n) \to H^2_k$ is causal, which means (see Proposition 4.1) that

$$W \cong [W^c_{0,\ldots,0}(z), zW^c_{0,\ldots,1}(z), \ldots, zW^c_{1,\ldots,1,0}(z), z^2 W^c_{0,\ldots,2}(z), \ldots],$$

for some

$$\{W^c_{i_1,\ldots,i_n}(z) \in H^\infty_{k\times K} : (i_1,\ldots,i_n) \in I_n\}.$$

Motivated by the above discussion, for $W : \ell^2(H^2_K) \to H^2_k$ time-invariant and causal, we introduce the operator

$$\begin{aligned}
W_c &\cong [W^c_{0,\ldots,0}(z), W^c_{0,\ldots,1}(z), \ldots, W^c_{1,\ldots,1,0}(z), W^c_{0,\ldots,2}(z), \ldots]\\
&= [W_{0,\ldots,0}(z), W_{0,\ldots,1}(z)/z, \ldots, W_{1,\ldots,1,0}(z)/z, W_{0,\ldots,2}(z)/z^2, \ldots].
\end{aligned} \tag{21}$$

We conclude this section by noting that to solve the (COP) problem (5), we can equivalently solve the following problem: Given $W : \ell^2(H_K^2) \to H_k^2$ time-invariant and causal as above, find

$$\sigma = \inf\{\|W - \Theta Q\| : QS_\Phi = UQ,\ Q \text{ causal}\}. \tag{22}$$

Thus we must solve the optimization problem (COP) on the Fourier transformed operators. This we will show how to explicitly do via a reduction theorem in the next section.

5. Reduction theorem. In this section, we formulate and prove our main result which will allow us to reduce the computation of a causal dilation to an ordinary one based on the classical commutant lifting theorem, i.e., interpolation in H^∞. In what follows $\mathcal{H}$, $\mathcal{K}$, $\mathcal{H}_i, i \geq 1$ will denote (complex, separable) Hilbert spaces.

To prove the result we will need two elementary lemmas.

LEMMA 5.1. *Let $A : \mathcal{K} \to \mathcal{H}$ be a bounded linear operator, and let T and S^* be isometries on $\mathcal{H}$ and $\mathcal{K}$, respectively. Then*

$$\|TAS\| = \|A\|.$$

Proof. By hypothesis, $T^*T = I$, and $SS^* = I$, and so

$$\begin{aligned}\|A\|^2 &= \|A^*A\| = \|A^*T^*TA\| \\ &= \|(TA)(TA)^*\| = \|TASS^*(TA)^*\| \\ &= \|(TAS)(TAS)^*\| = \|TAS\|^2,\end{aligned}$$

as required. □

LEMMA 5.2. *Let*

$$A = [A_1, A_2, \ldots] : \bigoplus_{i=1}^{\infty} \mathcal{H}_i \to \mathcal{H},$$

where

$$A(\oplus_{i=1}^{\infty} h_i) := \sum_{i=1}^{\infty} A_i h_i.$$

Further, let U_i^* be an isometry on $\mathcal{H}_i$ for $i \geq 1$. Then

$$\|A\| = \|[A_1, A_2, \ldots]\| = \|[A_1U_1, A_2U_2, \ldots]\|.$$

Proof. Note that

$$[A_1U_1, A_2U_2, \ldots, A_nU_n, \ldots] = [A_1, A_2, \ldots, A_n, \ldots] \begin{bmatrix} U_1 & 0 & 0 & \cdots & \cdots \\ 0 & U_2 & 0 & \cdots & \cdots \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & \ddots & U_n & \vdots \\ \vdots & \cdots & \cdots & \ddots & \ddots \end{bmatrix}.$$

However, if we set $S := \oplus_{i=1}^{\infty} U_i$, by hypothesis S^* is an isometry on $\oplus_{i=1}^{\infty} \mathcal{H}_i$, and so by Lemma 5.1, we are done. □

THEOREM 5.3 (Reduction theorem). *Notation as above. Then*

$$
\begin{aligned}
&(23)\quad \sigma = \inf\{\|W - \Theta Q\|\} : QS = UQ,\ Q \text{ causal}\}\\
&(24)\quad \quad = \inf\{\|[W_{0,\dots,0}(z) - \Theta q_{0,\dots,0}(z), z(W_{0,\dots,1}(z) - \Theta q_{0,\dots,1}(z)),\dots]\| :\\
&\qquad\qquad [q_{0,\dots,0}(z),\dots,q_{i_1,\dots,i_n}(z),\dots] \in \mathcal{L}(\ell^2(H^2_K), H^2_k),\ (i_1,\dots,i_n) \in I_n\}\\
&(25)\quad \quad = \inf\{\|W_c - \Theta Q\| : QS = UQ\}.
\end{aligned}
$$

(*Note in* (24) *the norm is the operator norm in* $\mathcal{L}(\ell^2(H^2_K), H^2_k)$. *In general, for Hilbert spaces* $\mathcal{H}$ *and* $\mathcal{K}$, $\mathcal{L}(\mathcal{H},\mathcal{K})$ *denotes the space of bounded linear operators from* $\mathcal{H}$ *to* $\mathcal{K}$.)

Proof. The second equality (24) follows from Proposition 4.1. To prove the third equality (25), it is enough to prove that for any causal, time-invariant operator

$$\Omega \cong [\omega_{0,\dots,0}(z), \omega_{0,\dots,1}(z), \dots, \omega_{i_1,\dots,i_n}(z), \dots],$$

we have $\|\Omega\| = \|\Omega_c\|$. (See (21) above.)

Now since

$$
\begin{aligned}
\|\Omega\| &= \text{ess sup}\{\|[\omega_{0,\dots,0}(\zeta), \omega_{0,\dots,1}(\zeta),\dots,\omega_{i_1,\dots,i_n}(\zeta),\dots]\| : |\zeta| = 1\},\\
\|\Omega_c\| &= \text{ess sup}\{\|[\omega_{0,\dots,0}(\zeta), \omega^c_{0,\dots,1}(\zeta),\dots,\omega^c_{i_1,\dots,i_n}(\zeta),\dots]\| : |\zeta| = 1\},
\end{aligned}
$$

we need to prove that for any fixed $\zeta \in \partial D$ that

$$\|[\omega_{0,\dots,0}(\zeta), \omega_{0,\dots,1}(\zeta),\dots,\omega_{i_1,\dots,i_n}(\zeta),\dots]\| = \|[\omega_{0,\dots,0}(\zeta), \omega^c_{0,\dots,1}(\zeta),\dots,\omega^c_{i_1,\dots,i_n}(\zeta),\dots]\|.$$

However, by Proposition 4.1,

$$\omega_{i_1,\dots,i_n}(\zeta) = \omega^c_{i_1,\dots,i_n}(\zeta)\zeta^{\max\{i_1,\dots,i_n\}} I_{\mathbf{C}^K},$$

where $I_{\mathbf{C}^K}$ is the identity on $\mathbf{C}^K$. Hence by Lemma 5.2 with $\mathcal{H}_i := \mathbf{C}^K$ and $U_i := \zeta^{\max\{i_1,\dots,i_n\}} I_{\mathbf{C}^K}$ $(i \geq 1)$, we are done. □

6. Algorithm for computation of σ. We would like to summarize the above discussion with a high-level algorithm for the computation of the optimal causal performance σ, and corresponding causal optimal interpolant Q_{opt} in (5) and (6).

First, using the notation of Theorem 5.3, let us denote

$$\sigma_o := \inf\{\|W_c - \Theta Q\| : QS = UQ\}. \tag{26}$$

(See equation (25).) Then Theorem 5.3 guarantees that

$$\sigma = \sigma_o.$$

This means that a causal optimization problem can be reduced to a classical generalized interpolation problem in H^∞.

We can summarize the procedure as follows:

(i) Let W, Θ be as in (5). (Thus $W : H^2_K(D^n) \to H^2_k$ here.) We compute $W(z_1^{i_1}\cdots z_n^{i_n})$ where $(i_1,\dots,i_n) \in I_n$. By (18) and (19), we get

$$W\Phi^* \cong [W_{0,\dots,0}(z), W_{0,\dots,1}(z),\dots,W_{i_1,\dots,i_n}(z),\dots],$$

and then by (20) we obtain the row matrix

$$[W_{0,\dots,0}(z), W^c_{0,\dots,1}(z),\dots,W^c_{i_1,\dots,i_n}(z),\dots].$$

(ii) The row matrix represents an operator (see (17)) $W_c : \ell^2(H^2_K) \to H^2_k$. Let $\Pi : H^2_k \to H^2_k \ominus \Theta H^2_k$ denote orthogonal projection. Using skew Toeplitz theory ([8], [17], [20]), we can compute the norm of the operator

$$\Lambda(W,\Theta) := \Pi W_c. \tag{27}$$

This norm is σ, the optimal causal performance.

(iii) Using the classical commutant lifting theorem and skew Toeplitz theory, we can compute the optimal dilation $B_c : \ell^2(H^2_K) \to H^2_k$ of $\Lambda(W,\Theta)$. Recall this means that

$$B_c S_\Phi = U B_c, \quad \Pi B_c = \Lambda(W,\Theta), \quad \|B_c\| = \|\Lambda(W,\Theta)\| = \sigma.$$

We can then write

$$B_c = W_c - \Theta Q_{opt,c}.$$

Then from (21), we can find the *optimal causal dilation*

$$B = W\Phi^* - \Theta Q_{opt}\Phi^*.$$

Note that B and B_c are related as in (21), and similarly for $Q_{\mathrm{opt},c}$ and $Q_{\mathrm{opt}}\Phi^*$. $Q_{\mathrm{opt}} : H^2_K(D^n) \to H^2_k$ is the *optimal causal interpolant*, i.e.,

$$\sigma = \|W - \Theta Q_{\mathrm{opt}}\|.$$

In the next section, we will give an explicit procedure for the computation of Q_{opt} in the SISO case.

7. Maximal vectors and optimal dilations. We use the notation of the previous section. We want to show how to compute the optimal dilation for

$$A := \Lambda(W,\Theta) : \ell^2(H^2) \to H^2.$$

(We are only considering SISO systems here.)

Our discussion will be based on [15], which generalizes a well-known result of Sarason [24]. We recall that a *maximal vector* of A, $h_o \neq 0$, is a vector such that $\|Ah_o\| = \|A\|\|h_o\|$.

Given $h \in \ell^2(H^2)$,

$$h = [h_1, h_2, \ldots]',$$

we write

$$h^* = [\bar{h}_1, \bar{h}_2, \ldots].$$

Moreover, we set

$$T := \Pi U | H^2 \ominus \Theta H^2,$$

where $\Pi : H^2 \to H^2 \ominus \Theta H^2$ denotes orthogonal projection. As above, U is the unilateral shift on H^2, and S_Φ denotes the shift on $\ell^2(H^2)$.

With this notation, we can now state the following result.

PROPOSITION 7.1. *Notation as above. Let $A : \ell^2(H^2) \to H^2 \ominus \Theta H^2$ be as above (so that $AU = TA$). Suppose moreover that that A has a maximal vector h_o. Let $B_c : \ell^2(H^2) \to H^2$ be the minimal intertwining dilation of A, i.e., $\Pi B_c = A$, $B_c U = S_\Phi B_c$, and $\|A\| = \|B_c\|$. Then if we let $\lambda := \|A\|^2$, we have that*

$$B_c = \frac{\lambda h_o^*}{Ah_o}.$$

Proof. We sketch the proof following [15]. First, given $h_o \in H$, we represent h_o as a column vector with components h_j, $j \geq 1$ as above. Let

$$B_c \cong [b_1, b_2, \ldots].$$

Then we have that

$$(B_c h_o)(z) = \sum_{j\geq 1} b_j(z) h_j(z)$$

(for $z \in D$), and

$$\|B_c\| = \text{ess sup}\left\{ \left(\sum_{j=1}^{\infty} |b_j(\zeta)|^2 \right)^{1/2} : |\zeta| = 1 \right\}.$$

However,

$$\|A\|^2\|h_o\|^2 = \|Ah_o\|^2 \leq \|B_c h_o\|^2 \leq \|B_c\|^2\|h_o\|^2 = \|A\|^2\|h_o\|^2.$$

Thus $\|Ah_o\|^2 = \|B_c h_o\|^2$, and since $\Pi B_c h_o = Ah_o$, we have that $Ah_o = B_c h_o$. Next note that

$$\sum_{j\geq 1} |b_j(e^{it})|^2 \leq \lambda$$

almost everywhere, and

$$\frac{1}{2\pi}\int_0^{2\pi} \left(\lambda \sum_{j=1}^{\infty} |h_j(e^{it})|^2 - |\sum_{j=1}^{\infty} b_j(e^{it}) h_j(e^{it})|^2 \right) dt = 0.$$

(This follows from the fact that $\lambda\|h_o\|^2 = \|B_c h_o\|^2$.) Using the Cauchy–Schwarz inequality, the expression under the integral sign is nonnegative. Thus

$$\begin{aligned} \lambda \sum_{j\geq 1} |h_j(e^{it})|^2 &= |\sum_{j\geq 1} b_j(e^{it}) h_j(e^{it})|^2 \\ &\leq \left(\sum_{j\geq 1} |b_j(e^{it})|^2 \right)\left(\sum_{j\geq 1} |h_j(e^{it})|^2 \right) \leq \lambda \sum_{j\geq 1} |h_j(e^{it})|^2 \end{aligned}$$

almost everywhere, which implies that

$$\sum_{j\geq 1} |b_j(e^{it})|^2 = \lambda$$

almost everywhere, and

$$h_j = \phi(e^{it})\overline{b_j(e^{it})}$$

almost everywhere for all $j \geq 1$, and for some function $\phi \in H^2$ satisfying

$$Ah_o = B_c h_o = \lambda\phi.$$

Thus for

$$B_c(e^{it}) \cong [b_1(e^{it}), b_2(e^{it}), \ldots]$$

we have

$$B_c(e^{it})\overline{Ah_o(e^{it})} = \lambda h_o(e^{it})^*$$

almost everywhere, as required. □

Remarks. (i) As remarked above, from the optimal dilation B_c, we can solve for $Q_{\text{opt},c}$ via

$$B_c = W_c - \Theta Q_{\text{opt},c}.$$

The optimal causal interpolant is then derived as described as in the last section.

(ii) In some cases it may be more convenient to derive the optimal dilation from a maximal vector of A^*. A similar proof to the one just given shows that

$$B_c = \frac{\overline{A^* h_1}}{\overline{h_1}}, \tag{28}$$

where $h_1 \in H^2 \ominus \Theta H^2$ is a maximal vector for A^*.

8. Nonlinear Control Problem. We will now describe the physical control problem in which we are interested. In our treatment that follows, we will add the causality constraint to the results of [15], and thereby derive a physically realizable nonlinear optimization procedure. First, we will need to consider the precise kind of input/output operator we will be considering. As above, H^2_k denotes the standard Hardy space of $\mathbf{C}^k$-valued functions on the unit disc. We now make the following definition.

Then we say an analytic input/output operator $\phi : H^2_k \to H^2_k$ is *admissible* if it is causal, time-invariant, majorizable, and $\phi(0) = 0$. We denote

$$C_l := \{\text{space of admissible operators}\}.$$

Since the theory we are considering is local, the notion of admissibility is sufficient for all of the applications we have in mind.

We now begin to formulate our control problem. Referring to Fig. 1, P represents a physical plant that we assume is modeled by an admissible operator. In our problem, we are required to design a feedback compensator C in such a way as to attentuate the effect of the filtered disturbances (filtered by the "weight" W) d. The unfiltered disturbances v are assumed to have energy (i.e., 2-norm) bounded by some fixed constant. This leads to following kind of mathematical problem. See [14] and [15] for more details.

Let P, W denote admissible operators, with W invertible. Then we say that the feedback compensator C *stabilizes* the closed loop if the operators $(I + P \circ C)^{-1}$ and

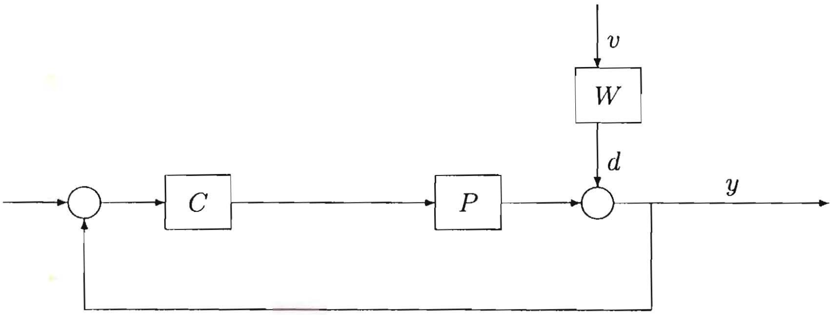

FIG. 1. *Standard feedback configuration.*

$C \circ (I + P \circ C)^{-1}$ are well defined and admissible. We can show that C stabilizes the closed loop if and only if

$$C = \hat{q} \circ (I - P \circ \hat{q})^{-1} \tag{29}$$

for some $\hat{q} \in \mathcal{C}_l$. (See [14], [15] and the references therein.) Note that the *weighted sensitivity* $(I + P \circ C)^{-1} \circ W$ can be written as $W - P \circ q$, where $q := \hat{q} \circ W$. This is precisely the operator relating the disturbance v to the output y. (Since W is invertible, the data q and $\hat{q}$ are equivalent.) In this context, we will call such a q, a *compensating parameter.* Note that from the compensating parameter q, we get a stabilizing compensator C via the formula (29).

As in [15], the problem we would like to solve here is a nonlinear version of the classical disturbance attenuation problem. This corresponds to the "minimization" of the "sensitivity" $W - P \circ q$ taken over all admissible q. To formulate a precise mathematical problem, we need to say in what sense we want to minimize $W - P \circ q$. This we will do in the next section, where we will propose a notion of "sensitivity minimization" which seems quite natural to analytic input/output operators. For the linear case of sensitivity minimization see [10], [18] and the references therein.

9. Nonlinear sensitivity function. This section follows very closely the set-up of [15]. However, now we explicitly put in the causality constraint.

We begin by defining a fundamental object, namely a nonlinear version of *sensitivity.* We should note that while the optimal H^∞ measure of performance is a real number in the linear case [18], the measure of performance that seems to be more natural in this nonlinear setting is a certain function defined in a real interval. This new kind of performance criterion is one of the keys concepts developed in [14] and [15]. See also §11 for a further analysis of the physical meaning of our nonlinear sensitivity function.

To define our notion of sensitivity, we will first have to partially order germs of analytic mappings. All of the input/output operators here will be admissible. We also follow here our convention that for given $\phi \in \mathcal{C}_l$, ϕ_n will denote the bounded linear map on the space $(H_k^2)^{\otimes n} \cong H_K^2(D^n)$ (with $K = k^n$) associated to the n-linear part of ϕ, which we also denote by ϕ_n (and which we always assume without loss of generality is symmetric in its arguments). The context will always make the meaning of ϕ_n clear.

We can now state the following definition.

DEFINITION 2. (i) *For* $W, P, q \in \mathcal{C}_l$ *(*W *is the weight,* P *the plant, and* q *the compensating parameter), we define the* sensitivity function $S(q)$,

$$S(q)(\rho) := \sum_{n=1} \rho^n \|(W - P \circ q)_n\|$$

for all $\rho > 0$ *such that the sum converges. Note that for fixed* P *and* W*, for each* $q \in \mathcal{C}_l$*, we get an associated sensitivity function.*

(ii) *We write* $S(q) \preceq S(\tilde{q})$*, if there exists a* $\rho_o > 0$ *such that* $S(q)(\rho) \leq S(\tilde{q})(\rho)$ *for all* $\rho \in [0, \rho_o]$*. If* $S(q) \preceq S(\tilde{q})$ *and* $S(\tilde{q}) \preceq S(q)$*, we write* $S(q) \cong S(\tilde{q})$*. This means that* $S(q)(\rho) = S(\tilde{q})(\rho)$ *for all* $\rho > 0$ *sufficiently small, i.e.,* $S(q)$ *and* $S(\tilde{q})$ *are equal as germs of functions.*

(iii) *If* $S(q) \preceq S(\tilde{q})$*, but* $S(\tilde{q}) \not\preceq S(q)$*, we will say that* q ameliorates $\tilde{q}$*. Note that this means* $S(q)(\rho) < S(\tilde{q})(\rho)$ *for all* $\rho > 0$ *sufficiently small.*

Now with Definition 2, we can define a notion of "optimality" relative to the sensitivity function.

DEFINITION 3. (i) $q_o \in \mathcal{C}_l$ *is called* optimal *if* $S(q_o) \preceq S(q)$ *for all* $q \in \mathcal{C}_l$.

(ii) *We say* $q \in \mathcal{C}_l$ *is* optimal with respect to its nth term q_n*, if for every* n*-linear* $\hat{q}_n \in \mathcal{C}_l$*, we have*

$$S(q_1 + \cdots + q_{n-1} + q_n + q_{n+1} \ldots) \preceq S(q_1 + \cdots + q_{n-1} + \hat{q}_n + q_{n+1} + \cdots).$$

If $q \in \mathcal{C}_l$ *is optimal with respect to all of its terms, then we say that it is* partially optimal.

10. Iterative causal commutant lifting method. In this section, we discuss a construction from which we will derive both partially optimal and optimal compensators relative to the sensitivity function given in Definition 2 above. As before, P will denote the plant, and W the weighting operator, both of which we assume are admissible. We always suppose that P_1 (the linear part of P) is an isometry, i.e., P_1 is a $k \times k$ inner matrix-valued H^∞ function. (P_1 corresponds to Θ of §6.)

We begin by noting the following key relationship:

$$(W - P \circ q)_l = W_l - \sum_{1 \leq j \leq l} \sum_{i_1 + \cdots + i_j = l} P_j(q_{i_1} \otimes \cdots \otimes q_{i_j}) \quad \forall l \geq 1.$$

Note that once again for ϕ admissible, ϕ_n denotes the n-linear part of ϕ, as well as the associated linear operator on $H^2_K(D^n)$.

We are now ready to formulate the *iterative causal commutant lifting procedure.* Let $\Pi : H^2_k \to H^2_k \ominus P_1 H^2_k$ denote orthogonal projection. Using the above (see (27)) we may choose q_1 causal such that

$$\|W_1 - P_1 q_1\| = \|\Lambda(W_1, P_1)\|.$$

Now given this q_1, we choose a causal q_2 such that

$$\|W_2 - P_2(q_1 \otimes q_1) - P_1 q_2\| = \|\Lambda(W_2 - P_2(q_1 \otimes q_1)), P_1)\|.$$

Inductively, given $q_1, \ldots, q_{n-1}$, set

$$\hat{W}_n := \left(W_n - \sum_{2 \leq j \leq n} \sum_{i_1 + \cdots + i_j = n} P_j(q_{i_1} \otimes \cdots \otimes q_{i_j}) \right) \tag{30}$$

for $n \geq 2$. Then we may choose q_n such that

$$\|\hat{W}_n - P_1 q_n\| = \|\Lambda(\hat{W}_n, P_1)\|. \tag{31}$$

Note that in each step of the procedure, the new "weight" $\hat{W}_n$ is determined by the n-linear part W_n of the original weight, and the optimal causal parameters chosen previously (namely, $q_1, \ldots, q_{n-1}$). The "plant" P_1 remains fixed throughout the procedure. Thus if P_1 is rational, the iterative causal commutant lifting procedure takes place on the finite dimensional space $H_k^2 \ominus P_1 H_k^2$, and may therefore be reduced to *finite matrix computations.* This will be illustrated with an example in §13.

The following facts can be proven just as in [14] and [15], to which we refer the reader for the proofs. (See in particular [15, pp. 849–853].) First the causal iterative commutant lifting procedure converges:

PROPOSITION 10.1. *With the above notation, let $q^{(1)} := q_1 + q_2 + \cdots$. Then $q^{(1)} \in \mathcal{C}_l$.*

Next given any $q \in \mathcal{C}_l$, we can apply the causal iterative commutant lifting procedure to $W - P \circ q$. Now set

$$S_C(q)(\rho) := \sum_{n=1} \rho^n \|\Lambda(\hat{W}_n, P_1)\|.$$

Then we have the following result.

PROPOSITION 10.2. *Given $q \in \mathcal{C}_l$, there exists $\tilde{q} \in \mathcal{C}_l$, such that $S(\tilde{q}) \equiv S_C(q)$. Moreover $\tilde{q}$ may be derived from the causal iterated commutant lifting procedure.*

Moreover, as in [15] we have the following results.

PROPOSITION 10.3. *q is partially optimal if and only if $S(q) \cong S_C(q)$.*

THEOREM 10.4. *For given P and W as above, any $q \in \mathcal{C}_l$ is either partially optimal or can be ameliorated by a partially optimal compensating parameter.*

Finally we have the following result.

THEOREM 10.5. *Let P and W be single-input/single-output admissible operators. If the linear part of P is rational, then the partially optimal compensating parameter q_{opt} constructed by the iterated causal commutant lifting procedure is optimal.*

The proof of this last result is based on the uniqueness of the optimal interpolant in the case when $k = 1$, and when the space $H^2 \ominus P_1 H^2$ is finite-dimensional. In fact, the conclusion of Theorem 10.5 remains valid under the hypotheses that the operators ΠW_j, $j \geq 1$ and ΠP_i, $i \geq 2$ are compact (and $k = 1$). See [15].

11. Control interpretation of iterated lifting. We would like to mention here what we believe to be a very natural way of looking at the optimization procedure discussed above. For convenience, we will only treat SISO systems here.

We refer again to Fig. 1. We consider the problem of finding

$$\mu_\delta := \inf_C \sup_{\|v\| \leq \delta} \|[(I + P \circ C)^{-1} \circ W]v\|, \tag{32}$$

where we assume all the operators involved are admissible. Thus we are looking at a worst case disturbance attenuation problem where the energy of the signals v is required to be bounded by some prespecified level δ. (Of course in the linear case since everything scales, we can always without loss of generality take $\delta = 1$. For nonlinear systems, we must specify the energy bound a priori.) Again with the assumptions made in §8, we see that (32) is equivalent to the problem of finding

$$\mu_\delta = \inf_{q \in \mathcal{C}_l} \sup_{\|v\| \leq \delta} \|(W - P \circ q)v\|. \tag{33}$$

The iterated causal commutant lifting procedure gives an approach for approximating a solution to such a problem. Briefly, the idea is that we write

$$\begin{aligned} W &= W_1 + W_2 + \cdots, \\ P &= P_1 + P_2 + \cdots, \\ q &= q_1 + q_2 + \cdots, \end{aligned}$$

where W_j, P_j, q_j are homogeneous polynomials of degree j. Note that

$$\mu_\delta = \delta \inf_{q_1 \in H^\infty} \|W_1 - P_1 q_1\| + O(\delta^2), \tag{34}$$

where the latter norm is the operator norm (i.e., H^∞ norm). From the classical commutant lifting theorem we can find an optimal (linear, causal, time-invariant) $q_{1,\text{opt}} \in H^\infty$ such that

$$\mu_\delta = \delta \|W_1 - P_1 q_{1,\text{opt}}\| + O(\delta^2). \tag{35}$$

Now the iterative procedure gives a way of giving higher-order corrections to this linearization. Let us illustrate this now with the second-order correction. Indeed, having fixed now the linear part $q_{1,\text{opt}}$ of q in (33), we note that

$$\begin{aligned} &W(v) - P(q(v)) - (W_1 - P_1 q_{1,\text{opt}})(v) \\ &\qquad = W_2(v) - P_2(q_{1,\text{opt}}(v)) - P_1 q_2(v) + \text{higher-order terms.} \end{aligned}$$

Regarding $\hat{W}_2, P_2, q_2$ as linear operators on $H^2 \otimes H^2 \cong H^2(D^2, \mathbf{C})$ as above, we see that

$$\sup_{\|v\| \le \delta} \|(W - P \circ q)(v) - (W_1 - P_1 q_{1,\text{opt}})v\| \le \delta^2 \|\hat{W}_2 - P_1 q_2\| + O(\delta^3),$$

where the "weight" $\hat{W}_2$ is given as in (30). The point of the iterative causal commutant lifting procedure is now to pick an optimal admissible $q_{2,\text{opt}}$, and so on.

In short, instead of simply designing a linear compensator for a linearization of the given nonlinear system, this methodology allows us to explicitly take into account the higher-order terms of the nonlinear plant, and therefore increase the ball of operation for the nonlinear controller.

12. Rationality. A nice feature of the iterated procedure described above is that if we start out with rational data, then we derive compensating parameters at each step that are also rational. Thus the whole procedure is amenable to digitable implementation in such cases. Let us briefly review the notion of rationality in this context. See [14] for all the details.

Let $W : H^2_K(D^n) \to H^2_k$ be time-invariant and admit the row vector representation

$$W\Phi^* \cong [W_{0,\ldots,0}(z), W_{0,\ldots,1}(z), \ldots, W_{i_1,\ldots,i_n}(z), \ldots],\ (i_1, \ldots, i_n) \in I_n.$$

Then we say that W is *rational* if there exists a numerical polynomial $q(z) \neq 0$ such that

$$q(z)[W_{0,\ldots,0}(z), W_{0,\ldots,1}(z), \ldots, W_{i_1,\ldots,i_n}(z), \ldots]$$

is a row of matrix-valued polynomials of bounded degree. Moreover if W is causal, we say that W is *causal rational* if

$$W_c \cong [W_{0,\ldots,0}(z), W^c_{0,\ldots,1}(z), \ldots, W^c_{i_1,\ldots,i_n}(z), \ldots]$$

is rational in the above sense.

The following result may be derived exactly as in [15, (see Thm. 8.7)].

THEOREM 12.1. *Notation as above. Suppose that the linear part of the plant is rational. Then the class of causal rational input/output operators is preserved under the causal iterated commutant lifting procedure.*

Hence for this important class of systems, we are reduced to rational finite-dimensional operations in carrying out our optimization procedure.

13. Example. In this section, we will give an example of our nonlinear design procedure. In what follows below, we set $H_{D^2} := H^2(D^2)$, the space of $\mathbf{C}$-valued analytic functions on the bidisc D^2 with square integrable boundary values. We should note that this example was first worked in [15] without the causality constraint that we impose now.

We let

$$W(z) = \frac{1-z}{2}$$

and $P = P_1 + P_2$ where P_1 is the operator given by multipication by z^2 (in the discrete Fourier domain), and

$$P_2(F) = \frac{1}{2\pi i}\int_{|\zeta|=1} F(z\zeta^{-1}, \zeta)\frac{d\zeta}{\zeta}$$

for $F \in H_{D^2} \cong H^2 \otimes H^2$. More precisely, as we explained above, we can regard a bilinear map P_2 on $H^2 \times H^2$ as a linear map on $H^2 \otimes H^2$, and then we identify $H^2 \otimes H^2$ with H_{D^2}. (The identification is given by $z \otimes 1 \to z_1$ and $1 \otimes z \to z_2$.) Notice that in the discrete-time domain, P_2 is just discrete Fourier transform of the "squaring" map, i.e., given the square integrable sequence $\{a_n\}$, we have that P_2 is the Fourier transform of the mapping $\{a_n\} \to \{a_n^2\}$. Thus it is clear that P_2 is causal.

We now apply our procedure to the weight W and the plant P. By slight abuse of notation, we let $W : H^2 \to H^2$ denote the operator defined by multiplication by W, and let $\Pi : H^2 \to H^2 \ominus P_1 H^2 =: H_1$ be orthogonal projection. We set $A_o := \Pi W | H_1$. Note that $H_1 \cong \mathbf{C}^2$, and that via this isomorphism, we have the identification

$$A_o = \begin{bmatrix} \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

However,

$$A_o^* A_o = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} \end{bmatrix},$$

from which we get that $\|A_o\| = (\sqrt{5}+1)/2$, and that a maximal vector h_o (i.e., a vector such that $\|A_o h_o\| = \|A_o\|\|h_o\| \neq 0$) is given by

$$h_o := \begin{bmatrix} 1 \\ -\beta \end{bmatrix},$$

where $\beta := (\sqrt{5}-1)/2$. Using then the Sarason formula [24], we can compute that the optimal compensating parameter is

$$q_1 := \frac{\beta}{2(1-\beta z)}.$$

Of course, the above computation was based on standard linear H^∞-optimization theory. We now want to show how to get the optimal *causal* second-order compensating parameter.

For $F \in H_{D^2}$, let

$$F(z_1, z_2) = \sum_{j,k=0}^{\infty} F_{jk} z_1^j z_2^k.$$

Note that the action of the operator (see (30))

$$-\hat{W}_2 := \frac{4}{\beta^2} P_2(q_1 \otimes q_1)$$

on F is determined by its action on

$$F_{00} + \sum_{j=1}^{\infty} F_{j0} z_1^j + \sum_{k=1}^{\infty} F_{0k} z_2^k.$$

Thus to compute the row vector representing $-\hat{W}_2$, we need only compute

$$\begin{aligned}
&(-\hat{W}_2)(F_{00} + \sum_{j=1}^{\infty} F_{j0} z_1^j + \sum_{k=1}^{\infty} F_{0k} z_2^k) \\
&= \frac{1}{2\pi} \int_{|\zeta|=1} (\sum_{m\geq 0} \beta^m z^m \zeta^{-m})(\sum_{n\geq 0} \beta^n \zeta^n)(\sum_{\min\{j,k\}=0} F_{jk} z^j \zeta^{k-j}) \frac{d\zeta}{\zeta} \\
&= \sum_{\min\{j,k\}=0} (F_{jk} (\beta z)^{\max\{j,k\}})/(1-\beta^2 z).
\end{aligned}$$

We identify as above an operator $\Omega : H_K^2(D^n) \to H_k^2$ and its Fourier transformed version $\Omega\Phi^* : \ell^2(H_K^2) \to H_k^2$.

Therefore (under this identification),

$$-\hat{W}_2 \cong \frac{1}{1-\beta^2 z}[1, \beta z, \beta z, \beta^2 z^2, \ldots, \beta^n z^n, \ldots],$$
$$-\hat{W}_{2,c} \cong \frac{1}{1-\beta^2 z}[1, \beta, \beta, \beta^2, \ldots, \beta^n, \ldots],$$

and

$$\|\hat{W}_2\| = \|\hat{W}_{2,c}\| \approx 2.4195.$$

Set $A = \Pi(-\hat{W}_{2,c})$, where $\Pi : H^2 \to H^2 \ominus z^2 H^2 =: H(z^2) \cong \mathbf{C}^2$ denotes orthogonal projection. Note that the compressed shift T on $H(z^2) \cong \mathbf{C}^2$ is given by the truncated Toeplitz operator

$$T = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Using skew Toeplitz theory ([8], [17], [20]), we compute the norm of A and the corresponding optimal vector. Accordingly, we let $r(z) := 1 - \beta^2 z$. Then for $\rho > 0$, and for

$$\lambda := \frac{2-\beta}{(2\beta-1)\rho^2},$$

we compute that

$$\begin{aligned} r(T)(\rho^2 I_{\mathbf{C}^2} - AA^*)r(T)^* &= \rho^2 r(T)r(T)^* - (1 + 2\sum_{i=1}^{\infty}\beta^{2i})I_{\mathbf{C}^2} \\ &= (1-\beta)\rho^2 \begin{bmatrix} 1 + 1/\beta - \lambda & -1 \\ -1 & 3-\lambda \end{bmatrix}. \end{aligned}$$

$\|A\|$ is given by the largest ρ such that the latter matrix is singular. Thus we see that

$$\|\Pi(-\hat{W}_{2,c})\| = \|A\| \approx 1.8079,$$

which is the optimal *causal* performance. If we drop the causality requirement, then we get that

$$\|\Pi(-\hat{W}_2)\| \approx 1.4314.$$

(Of course, with the additional constraint the norm of the optimal dilation increases.)

Let

$$y_o(z) := 1 + \left(1 + \frac{1}{\beta} - \lambda\right) z \in H(z^2),$$

so that we may regard

$$y_o = \begin{bmatrix} 1 \\ 1 + \frac{1}{\beta} - \lambda \end{bmatrix}$$

under the identification $H(z^2) \cong \mathbf{C}^2$. Then it is easy to compute that

$$r(T)(\|A\|^2 I_{\mathbf{C}^2} - AA^*)r(T)^* y_o = 0.$$

Therefore $r(T)^* y_o$ is a maximal vector of A^*. But from the previous section (see (28)), the optimal dilation $B_{\mathrm{opt},c}$ of A is

$$\begin{aligned} B_{\mathrm{opt},c} &\cong \frac{\overline{A^* r(T)^* y_o}}{\overline{r(T)^* y_o}} \\ &= \frac{(3-\lambda)z + 1}{(1 + \frac{1}{\beta} - \lambda)z + 1}[1, \beta, \beta, \beta^2, \beta^2, \ldots]. \end{aligned}$$

Thus the optimal *causal* dilation B_{opt} of $\Pi(-\hat{W}_2)$ is

$$B_{\mathrm{opt}} \cong \frac{(3-\lambda)z + 1}{(1 + \frac{1}{\beta} - \lambda)z + 1}[1, \beta z, \beta z, \beta^2 z^2, \ldots].$$

The optimal causal interpolant q_2 is derived from

$$-\frac{4}{\beta^2} P_2(q_1 \otimes q_1) - z^2 q_2 = -B_{opt},$$

which gives that

$$q_2 \cong \frac{(\lambda - 3)\beta^2}{(1-\beta^2 z)((1+1/\beta-\lambda)z+1)}[1, \beta z, \beta z, \beta^2 z^2, \ldots].$$

Now set $q^{(2)} := q_1 + q_2$, the optimal second-order compensating parameter, and $\hat{q}^{(2)} := q^{(2)} W^{-1}$. The resulting controller is given by $C^{(2)} = \hat{q}^{(2)} \circ (I - P \circ \hat{q}^{(2)})^{-1}$. Note that it is not necessary to explicitly compute $C^{(2)}$, since it can be implemented in a feedback loop with components P and $\hat{q}^{(2)}$ as in [27].

14. Concluding remarks. In this paper, we have given an iterative approach for the construction of optimal causal compensators for input/output operators described by analytic mappings. Our procedure generalizes weighted sensitivity H^∞ minimization in a straightforward natural way. Hence, it may be regarded as a weighted nonlinear inversion procedure.

In contrast to our previous work using power series approaches ([3], [4], [14], [15]), we can now *guarantee causality* a priori. Moreover, the computation of a causal compensator can be reduced to classical dilation theory, and in fact the skew Toeplitz techniques of [8], [17], and [20] provide an explicit computational methodology.

The example which we have worked out here, has been given just for the purpose of illustrating our procedure. We plan to work out a more complicated and realistic problem, the details of which will be given in an upcoming report.

REFERENCES

[1] W. ARVESON, *Interpolation problems in nest algebras*, J. Funct. Anal., 20 (1975), pp. 208–233.

[2] M. ATIYAH AND I. MACDONALD, *Introduction to Commutative Algebra*, Addison-Wesley, New York, 1969.

[3] J. BALL, C. FOIAS, J. W. HELTON, AND A. TANNENBAUM, *On a local nonlinear commutant lifting theorem*, Indiana J. Math., 36 (1987), pp. 693–709.

[4] ———, *Nonlinear interpolation theory in H^∞*, in Modelling, Robustness, and Sensitivity in Control Systems, Ruth Curtain, ed., Springer-Verlag, New York, 1987.

[5] ———, *A Poincaré-Dulac approach to a nonlinear Beurling-Lax-Halmos theorem*, J. Math. Anal. Appl., 139 (1989), pp. 496–514.

[6] J. BALL AND J. W. HELTON, *Sensitivity bandwidth optimization for nonlinear feedback systems*, Tech. Report, Dept. of Math., Univ. of California at San Diego, 1988.

[7] ———, *H^∞ control for nonlinear plants: connections with differential games*, IEEE Conference on Decision and Control, Tampa, Florida, 1989, pp. 956–962.

[8] H. BERCOVICI, C. FOIAS, AND A. TANNENBAUM, *On skew Toeplitz operators*. I, Oper. Theory Adv. Appl., 29 (1988), pp. 21–44.

[9] S. BOYD AND L. CHUA, *Fading memory and the problem of approximating nonlinear operators with Volterra series*, IEEE Trans. Circuits Systems, CAS-32 (1985), pp. 1150–1161.

[10] J. DOYLE, B. FRANCIS, AND A. TANNENBAUM, *Feedback Control Theory*, MacMillan, New York, 1991.

[11] C. FOIAS, *Contractive intertwining dilations and waves in layered media*, Proc. Internat. Congress Mathematicians, Helsinki, 2 (1978), pp. 605–613.

[12] C. FOIAS AND A. FRAZHO, *The Commutant Lifting Approach to Interpolation Problems*, Birkhauser-Verlag, Boston, 1990.

[13] C. FOIAS, C. GU, AND A. TANNENBAUM, *Intertwining dilations, intertwining extensions and causality*, Acta Sci. Math. (Szeged), 57 (1993), pp. 101–123.

[14] C. FOIAS AND A. TANNENBAUM, *Iterated cummutant lifting for systems with rational symbol*, *Oper. Theory Adv. Appl.*, 41 (1989), pp. 255–277.

[15] ———, *Weighted optimization theory for nonlinear systems*, SIAM J. Control Optim., 27 (1989), pp. 842–860.

[16] ———, *Causality in commutant lifting theory*, J. Funct. Anal., 118 (1993), pp. 407–441.

[17] C. Foias, A. Tannenbaum, and G. Zames, *Some explicit formulae for the singular values of certain Hankel operators with factorizable symbol*, SIAM J. Math. Anal., 19 (1988), pp. 1081–1089.

[18] B. Francis, *A Course in H^∞ Control Theory*, McGraw-Hill, New York, 1981.

[19] B. Francis and A. Tannenbaum, *Generalized interpolation theory in control*, Math. Intelligencer, 10 (1988), pp. 48–58.

[20] C. Gu, *Eliminating the genericity conditions in the skew Toeplitz operator algorithm for H^∞ optimization*, SIAM J. Math. Anal., 23 (1992), pp. 1623–1636.

[21] E. Hille and R. Phillips, *Functional Analysis and Semigroups*, AMS Colloquium Publications, Vol. 31, American Mathematical Society, Providence, Rhode Island, 1957.

[22] J. W. Helton, *Broadbanding: gain equalization directly from data*, IEEE Trans. Circuits Systems, CAS-28 (1981), pp. 1125–1137.

[23] S. Parrott, *On the quotient norm and the Sz. Nagy-Foias lifting theorem*, J. Funct. Anal., 30 (1970), pp. 311–328.

[24] D. Sarason, *Generalized interpolation in H^∞*, Trans. Amer. Math. Soc., 127 (1967), pp. 179–203.

[25] B. Sz.-Nagy and C. Foias, *Harmonic Analysis of Operators on Hilbert Space*, North-Holland, Amsterdam, 1970.

[26] A. J. van der Schaft, *L_2 gain analysis of nonlinear systems and nonlinear state feedback H_∞ control*, IEEE Trans. Automat. Control, 37 (1992), pp. 770–784.

[27] G. Zames, *Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses*, IEEE Trans. Automat. Control, AC-26 (1981), pp. 301–320.